RHEUMATOLOGY

Randomized controlled trials and real-world data: differences and similarities to untangle literature data

Sara Monti[1],*, Vittorio Grosso[1],*, Monica Todoerti[1] and Roberto Caporali[1]

Abstract

Randomized controlled trials (RCTs) represent the gold-standard of medical evidence to assess the efficacy and safety of therapeutic interventions. However, the need to minimize bias and ensure the correct design to explore the study aims often affects the generalizability of results. As a consequence, the evidence derived from the most rigorous research strategy available is not always representative of the real-world settings for which this evidence is ultimately intended. Observational studies, in contrast, although affected by a number of potential confounders, can more effectively capture treatment characteristics and safety issues that had not been identified by previous RCTs, owing to the short duration of follow-up or highly selective inclusion criteria. The aim of this review is to provide a comparative summary of the main advantages and pitfalls of RCTs and real-world data, emphasizing the need for a constant integration of all available levels of evidence to provide the best care for patients.

Key words: randomized controlled trial, real-world data, study design, rheumatoid arthritis, biologic drugs

---

Rheumatology key messages

- Randomized controlled trials represent the best study design to assess efficacy of a therapeutic intervention.
- Results of randomized controlled trials often lack external validity (therefore generalizability) to the real-world population.
- Observational studies data need to be integrated, particularly to detect rare or late-onset adverse events.

REVIEW

Introduction

According to the principles of evidence-based medicine, randomized controlled trials (RCTs) represent the cornerstone of clinical research, providing the highest hierarchical level of evidence based on a single experiment [1]. Randomized controlled trials are pivotal in the development of new therapeutic strategies through the assessment of the efficacy of new drugs *vs* placebo, or the comparative performance against the standard of care. Among the advantages, the unbiased distribution of confounders, the minimized systematic differences in treatment allocation and the application of blinded procedures make RCTs the ideal study design in many fields of research. Nevertheless, RCTs are not immune to flaws. Indeed, in order to reduce the risk of bias, they require strict inclusion and exclusion criteria, thus limiting the generalizability of the findings to broader populations. Large sample sizes are usually required, making RCTs costly in terms of money and time. The high costs often impose short trial durations, which may not cover the detection of some rare or delayed side-effects and long-term efficacy.

Real-world data have become particularly relevant in filling some of the gaps left by data derived from the highly selected populations enrolled in RCTs. The main advantages of observational real-life studies are the better representation of routine clinical practice scenarios, the lower costs and the longer time of observation, thus optimizing the detection of adverse events (AEs). Observational data are extremely useful to improve the management of rare conditions, particularly in those cases still lacking generally recognized standards of care. Nonetheless, the risk of bias is much higher with observational studies. Although some statistical tools help to minimize the effect of confounders, not all sources of bias can be removed, making generalizability difficult to achieve even with this type of study design. An overview of the main differences between RCTs and observational studies is presented in Table 1.

[1]Department of Rheumatology, IRCCS Policlinico S. Matteo Foundation, University of Pavia, Pavia, Italy

Submitted 11 August 2017; revised version accepted 26 March 2018

*Sara Monti and Vittorio Grosso contributed equally to this study.

Correspondence to: Roberto Caporali, Department of Rheumatology, IRCCS Policlinico S. Matteo Foundation, University of Pavia, 27100 Pavia, Italy.
E-mail: caporali@smatteo.pv.it

TABLE 1 Main strengths and weaknesses of randomized controlled trials compared with observational studies

Type of study	Strengths	Weaknesses
Randomized clinical trials	Best for studying an intervention Randomized High internal validity Unbiased distribution of confounders Evaluates efficacy	Expensive: time and money Short follow-up Volunteer bias Low generalizability to different or real-world population
Observational cohorts (including national registries)	Best to study effects of risk factors on an outcome Participants can be matched Detect rare or late-onset adverse events Evaluate effectiveness	Channelling bias Exposure may be linked to a hidden confounder Blinding is difficult

In this review, we provide a comparative summary of the advantages and drawbacks RCTs and real-world data to promote a more critical interpretation of literature data, taking into account the advantages and pitfalls connected with the study design.

Patient selection criteria

Randomized controlled trials are designed to test the effects of an intervention under ideal, controlled circumstances.

Strict adherence to protocols, application of inclusion and exclusion criteria and patient randomization are all essential steps to reduce the risk of bias. The results of an RCT provide information regarding the expected effect of the intervention in a well-defined population and in a specific setting. The broader the population and the settings, the greater is the external validity of the study, which allows the conclusions to be generalized to routine clinical practice. On the contrary, RCTs are designed to maximize internal validity (which is the minimization of the possibility of bias regarding the effect of an intervention) at the expenses of generalizability [2].

As a result of strict selection criteria, patients enrolled in RCTs are usually younger, with fewer co-morbidities and shorter disease duration, and are therefore far from being representative of the everyday clinical practice patient. Therefore, the results of most RCTs, although strengthened by the rigorous methodology applied, do not allow the same conclusions to be drawn automatically in a different population or setting.

A review on the external validity of RCTs in the fields of cardiology, mental health and oncology found that a significant proportion of the general population suffering with those conditions would have not been included in RCTs, thus being prescribed the intervention with no direct evidence of the effects in that specific setting [3].

In the field of rheumatology, a clear example of the difficulties in transferring evidence from RCTs to the real world was demonstrated by Zink *et al.* [4] with the analysis of the German biologics register Rheumatoid Arthritis Observation of Biologic Therapy. The authors concluded that only 33% of the patients treated in real life with a TNF inhibitor (TNFi; infliximab, etanercept or adalimumab) would have been eligible for the major trials that had led

to the approval of the drugs. Therefore, high-quality evidence supporting the prescription of TNFi was lacking for about two-thirds of RA patients included in the German registry. Another example of this discrepancy was found when applying the eligibility criteria of 30 RCTs on biological drugs for RA to two observational clinical cohorts in the USA: the Veterans Affairs Rheumatoid Arthritis registry and the Rheumatology and Arthritis Investigational Network Database [5]. The authors concluded that only 3.7% of patients in the Veterans Affairs Rheumatoid Arthritis registry and 7.1% in the Rheumatology and Arthritis Investigational Network Database would have satisfied eligibility criteria for an RCT. Ineligibility was mostly explained by lower disease activity in the observational cohorts in comparison with RCTs. These findings can be explained, in part, by the design of RCTs used to support regulatory approval of new treatment agents, known as explanatory RTCs. Explanatory trials are specifically developed to demonstrate treatment efficacy and short-term safety. Typically, patients included in these trials have significantly higher levels of disease activity compared with clinical practice [5]. The other explanation, once more, is the effect of restrictive eligibility criteria not reflecting the characteristics of the everyday target population. A possible solution to improve the extrapolation of results to be applied in the routine care of patients with a specific disease is to use a pragmatic RCT design. While maintaining high internal validity, these trials more closely reflect clinical practice, allowing for a higher degree of heterogeneity of selected patients [6].

Efficacy and effectiveness

A relevant aspect that distinguishes data derived from RCTs and observational studies is the difference between efficacy and effectiveness of a specific intervention [7]. A treatment is effective when it produces the desired outcome in ideal circumstances; therefore, in the context of RCTs. Scientific evidence derived from RCTs clarifies what the expected outcomes of a single treatment are when offered to a well-defined population of patients with selected features [8]. The experimental nature of this process, including selection criteria, randomization and blinding applied to researchers and subjects,

provides results in artificial settings, which may not be fully extended to routine care. Consequently, most medical treatments are used in real life on patients groups and with doses and frequencies that have never been specifically studied in the ideal conditions of RCTs. Indeed, many patients assessed in routine care require multidrug treatments, with potential interactions, and may have multiple conditions influencing the outcome. Patients and disease characteristics may also differ significantly from the counterparts enrolled in clinical trials. These differences in the target population usually lead to discrepancies between RCTs and the real-world estimated magnitude of treatment effects [9].

Although the drug efficacy demonstrated by RCTs is not put into question, the effectiveness, defined as the extent to which a treatment achieves its intended effect in the usual clinical practice, might have not been investigated sufficiently before marketing approval. A treatment is effective if it works in real life in non-ideal circumstances, which are usually far from those observed in RCTs. To assess the effectiveness of a treatment, observational studies and real-world data in general are superior to RCTs, allowing for broader inclusion criteria and producing new evidence on how the drug performs in realistic clinical conditions. Observational studies ideally allow the inclusion of every case treated with the intervention, providing data on its outcomes in the complex environment of routine care.

In the case of RA, the effectiveness of treatments is usually lower when measured in real life compared with RCTs [10–12]. In particular, results from the Rheumatoid Arthritis DMARD Intervention and Utilization Study indicated no significant evidence of superiority in treating RA with infliximab + MTX *vs* MTX alone when the treatment was offered in routine clinical practice [12], in contrast to the results of a previously conducted RCT [13]. This conflicting evidence is explained by the difference in the study populations. Real-life patients often report higher lack of efficacy, loss of efficacy and AE rates compared with RCT patients. Possible explanations for this phenomenon are the higher prevalence of co-morbidities, lower baseline disease activity and lower adherence to treatment [14].

Although observational data are easily affected by bias, lacking randomization and control for confounders, they still represent the best available option to investigate the therapeutic effect of the intervention in routine clinical practice. Efficacy and effectiveness data should therefore be integrated and interpreted in combination.

Safety

The assessment of safety is another key element that can be influenced by the study design. In RCTs, small samples, strict patient selection criteria and short-term follow-up do not always allow the measurement, with sufficient statistical power, of the probability of rare AEs. Moreover, AE rates recorded during RCTs do not precisely predict the incidence of AEs in real-life settings. With the intent of reducing the risk of bias, the population enrolled in RCTs is selected with rigorous criteria and is blindly exposed to the drug for a limited period of time. These limits may lead

to an incorrect estimate of the risk for the single AE, which may result in an over- or underestimation. The latter was the case in the risk for statin-induced myopathy: although in RCTs ∼3% of patients developed myalgia, this condition was reported by 13% of the users in prospective clinical studies [15]. The 4-fold increase in the prevalence could result from several causes. The short duration of RCTs might not have allowed the event to be detected in patients who would have developed the AE after a longer period of drug exposure; the selection criteria might have excluded older patients and, possibly, other categories with higher risk for myopathy; the frequency of this AE reported from RCTs might have been lower by chance, and the relatively small sample size might not have allowed for a more precise estimate.

Another example of underestimation of risk in RCTs in represented by hypoglycaemia event rates in patients with type 1 diabetes mellitus and insulin-treated type 2 diabetes mellitus. A recent review compared the event rates between RCTs and real-world data. Although a large degree of overlap between the two settings existed, hypoglycaemia was generally more frequent in real-world populations [16]. The authors concluded that this difference might be explained by an underestimation of the event in RCTs, which usually exclude some populations with higher risk for hypoglycaemia, such as patients with renal dysfunction, elderly frail patients or those with malnutrition or co-morbidities. Moreover, patients enrolled in RCTs are usually subjected to more intensive monitoring than what is usually provided in routine clinical practice.

An example of risk overestimation is the issue of the development of cancer in patients treated with TNFi. When TNFi were prescribed to patients with RA for the first time, the results of RCTs suggested an increase in the risk of lymphoma [17, 18], and consequently, this AE was reported on the product label. However, observational real-world data derived from very large cohorts included in several registries found no evidence of an increased risk among patients assessed in routine clinical practice. The apparently higher risk can be explained by the channelling bias of treating those patients with the highest burden of disease-related inflammation, and therefore an increased risk of developing lymphoproliperative disorders, with TNFi. Therefore, the AE was not connected to the type of treatment, but rather, it was associated with the characteristics of the population [19, 20]. Likewise, another example of AE well known to the rheumatological community that had not emerged from the first RCTs in the early era of TNFi use is the reactivation of latent tuberculosis [21].

Once more, these examples underline how the risk/benefit ratio might be skewed if relying only on data from RCTs and that there is a constant need to integrate all levels of available evidence into the clinical management of patients.

Filling the gap between RCTs and real-world data

The issue of generalizability and extrapolation of the results of RCTs to the routine care of patients is a problem

that affects all medical disciplines. Results of RCTs are essential to inform international recommendations for specific conditions and to widen the therapeutic armamentarium. Nevertheless, continuous real-world data surveillance and analysis needs to be carried out to fill the efficacy–effectiveness gap and to capture fully the long-term safety of drugs when applied to the imperfect routine care patient. As advocated in the previous sections of this review, the treating clinician needs to combine all the available evidence, keeping in mind the advantages and limits related to the study design. Future research should be improved by a better integration of RCTs and observational studies, by applying selection criteria that better reflect the real-life target populations. Moreover, evidence synthesis tools should be implemented to offer a more critical point of view by combining all levels of available scientific evidence. Some of these tools are already well known. Meta-analyses offer the advantage of combining the results of several studies, increasing the sample size and the power of the results obtained. Another strategy that can be applied to improve the usability of the results obtained is to adapt the trial design according to its scope. Pragmatic RCTs can better reflect the conditions of routine practice. Adaptive trials can be adjusted in response to information generated during the trial, while n-of-1 trials or single subject clinical trials focus on individualized treatment strategies in the setting of high population heterogeneity [22].

Furthermore, in recent years, more structured efforts are being directed towards finding a way of providing scientific data with a high level of evidence, while being more applicable to the real target populations. The pragmatic-explanatory continuum indicator summary tool has been developed to allow trialists to design studies that better match with the needs of the intended users of the study results. Finally, this tool can also be used to assess the methodological quality of completed trials [3].

Conclusions

Randomized controlled trials remain the gold standard to assess the efficacy of an intervention. They are irreplaceable when the effect of a new treatment needs to be tested. Observational data can provide valuable evidence from real-life routine clinical care and identify previously unrecognized aspects related to treatment characteristics and safety.

All study designs, taking into account specific strengths and limitations, should contribute to generate the scientific evidence that will guide the real evidence-based approach to the patient.

References

1 Guyatt G, Oxman AD, Akl EA *et al*. GRADE guidelines: 1. Introduction—GRADE evidence profiles and summary of findings tables. J Clin Epidemiol 2011;64:383–94.

2 Rothwell PM. External validity of randomised controlled trials: ''to whom do the results of this trial apply?''. Lancet 2005;365:82–93.

3 Kennedy-Martin T, Curtis S, Faries D, Robinson S, Johnston J. A literature review on the representativeness of randomized controlled trial samples and implications for the external validity of trial results. Trials 2015;16:495.

4 Zink A, Strangfeld A, Schneider M *et al*. Effectiveness of tumor necrosis factor inhibitors in rheumatoid arthritis in an observational cohort study: comparison of patients according to their eligibility for major randomized clinical trials. Arthritis Rheum 2006;54:3399–407.

5 Vashisht P, Sayles H, Cannella AC, Mikuls TR, Michaud K. Generalizability of patients with rheumatoid arthritis in biologic agent clinical trials. Arthritis Care Res 2016;68:1478–88.

6 Nair SC, Kievit W, Janse RW *et al*. Generalization and extrapolation of treatment effects from clinical studies in rheumatoid arthritis. Arthritis Care Res 2014;66:998–1007.

7 Nordon C, Karcher H, Groenwold RHH *et al*. The ''efficacy-effectiveness gap'': historical background and current conceptualization. Value Health 2016;19:75–81.

8 Haynes B. Can it work? Does it work? Is it worth it? BMJ 1999;319:652–3.

9 Ioannidis JP, Haidich AB, Pappa M *et al*. Comparison of evidence of treatment effects in randomized and nonrandomized studies. JAMA 2001;286:821–30.

10 Martin K, Bentaberry F, Dumoulin C *et al*. Effectiveness and safety profile of leflunomide in rheumatoid arthritis: actual practice compared with clinical trials. Clin Exp Rheumatol 2005;23:80–4.

11 Gibofsky A, Palmer WR, Goldman JA *et al*. Real-world utilization of DMARDs and biologics in rheumatoid arthritis: the RADIUS (Rheumatoid Arthritis Disease-Modifying Anti-Rheumatic Drug Intervention and Utilization Study) study. Curr Med Res Opin 2006;22:169–83.

12 Weaver AL, Lautzenheiser RL, Schiff MH *et al*. Real-world effectiveness of select biologic and DMARD monotherapy and combination therapy in the treatment of rheumatoid arthritis: results from the RADIUS observational registry. Curr Med Res Opin 2006;22:185–98.

13 Blumenauer B, Judd M, Wells G *et al*. Infliximab for the treatment of rheumatoid arthritis. Cochrane Database Syst Rev 2002(3);CD003785.

14 Wolfe F, Michaud K. Towards an epidemiology of rheumatoid arthritis outcome with respect to treatment: randomized controlled trials overestimate treatment response and effectiveness. Rheumatology 2005;Suppl 4:iv18–22.

15 Sathasivam S, Lecky B. Statin induced myopathy. BMJ 2008;337:a2286

16 Elliott L, Fidler C, Ditchfield A, Stissing T. Hypoglycemia event rates: a comparison between real-world data and

randomized controlled trial populations in insulin-treated diabetes. Diabetes Ther 2016;7:45–60.

17 Askling J, Fahrbach K, Nordstrom B *et al*. Cancer risk with tumor necrosis factor alpha (TNF) inhibitors: meta-analysis of randomized controlled trials of adalimumab, etanercept, and infliximab using patient level data. Pharmacoepidemiol Drug Saf 2011;20:119–30.

18 Maini R, St Clair EW, Breedveld F *et al*. Infliximab (chimeric anti-tumour necrosis factor α monoclonal antibody) versus placebo in rheumatoid arthritis patients receiving concomitant methotrexate: a randomised phase III trial. Lancet 1999;354:1932–9.

19 Chen Y, Sun J, Yang Y, Huang Y, Liu G. Malignancy risk of anti-tumor necrosis factor alpha blockers: an overview of systematic reviews and meta-analyses. Clin Rheumatol 2016;35:1–18.

20 Mercer LK, Galloway JB, Lunt M *et al*. Risk of lymphoma in patients exposed to antitumour necrosis factor therapy: results from the British Society for Rheumatology Biologics Register for Rheumatoid Arthritis. Ann Rheum Dis 2017;76:497–503.

21 Keane J, Gershon S, Wise RP *et al*. Tuberculosis associated with infliximab, a tumor necrosis factor α-neutralizing agent. N Engl J Med 2001;345:1098–104.

22 Favalli EG, Bugatti S, Biggioggero M, Caporali R. Treatment comparison in rheumatoid arthritis: head-to-head trials and innovative study designs. Biomed Res Int 2014;2014:831603.